

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant: John Kroeker, et al.
Serial No.: 09/918,733
Filed: 31 July 2001
Title: METHOD OF AND SYSTEM FOR IMPROVING ACCURACY IN A
SPEECH RECOGNITION SYSTEM
Group Art Unit: 2626
Examiner: Azad, Abul K.
Atty. Docket No.: 57622-045 (ELZK-5)
Confirmation No.: 2704

CERTIFICATE OF MAILING OR TRANSMISSION

I hereby certify that this correspondence is being deposited with the U.S. Postal Service as first class mail in an envelope addressed to: Mail Stop After Final, Commissioner for Patents, P. O. Box 1450, Alexandria, VA 22313-1450, or facsimile transmitted (571-273-8300) to the USPTO, on the date indicated below.

Date: 08 August 2007


Erin Olson

Mail Stop After Final
Commissioner For Patents
P.O. Box 1450
Alexandria, VA 22313-1450

RESPONSE

This paper is in response to the final Office Action mailed 04 June 2007 in connection with the above-identified application.

The Applicants appreciate the Examiner's thorough examination of the subject application and request reconsideration and further examination in view of the following:

- **Listing of Claims**, beginning on Page 2 of this paper; and
- **Remarks**, beginning on Page 7 of this paper.

Applicants note that a Request for Continued Examiner (RCE) under 37 CFR § 1.114 is submitted with this paper.

Listing of Claims:

This listing of claims will replace all prior versions and listings of claims in the subject application.

Claims:

1. (Previously presented) A speech recognition system comprising:
 - a querying device for posing at least one query to a respondent over a telephone;
 - a speech recognition device which receives an audio response from said respondent over the telephone and conducts a speech recognition analysis of said audio response to automatically produce a corresponding text response;
 - a storage device for recording and storing said audio response as it is received by said speech recognition device;
 - an accuracy determination device for automatically comparing said text response to a text set of expected responses and determining whether said text response corresponds to one of said expected responses, wherein said accuracy determination device is configured and arranged to determine whether said text response corresponds to one of said expected responses within a predetermined accuracy confidence parameter and to flag said audio response so as to produce a flagged audio response for further review by a human operator when said text response does not correspond to one of said expected responses within said predetermined accuracy confidence parameter; and
 - a human interface device for enabling said human operator to hear said flagged audio response and review the corresponding text response for the flagged audio response to determine the actual text response for the flagged audio response, either by selecting from a pre-determined list of text responses or typing the actual text response if no such match exists in the pre-determined list of text responses.
- 2.-4. (Cancelled)
5. (Previously presented) The speech recognition system of claim 1, wherein said human interface device comprises a personal computer including a monitor for enabling the human operator to view said text responses and an audio speaker device for enabling the operator to listen to said

flagged audio responses.

6. (Previously presented) The speech recognition system of claim 5, wherein said querying device includes a program having an application file, said application file including code which causes the at least one query to be posed to the respondent, a list of expected responses and an address at which a file containing the received audio response will be stored in the storage device.

7. (Previously presented) The speech recognition system of claim 1, wherein said querying device includes a program having an application file, said application file including code which causes the at least one query to be posed to the respondent, a list of expected responses and an address at which a file containing the received audio response will be stored in the storage device.

8. (Previously presented) The speech recognition system of claim 1, wherein said human interface device includes a graphical user interface on which the human operator views said text set of expected responses, wherein after listening to said audio response, the human operator is able to select one of said expected responses from said text set of expected responses if the human operator determines that the response corresponds to one of said expected responses.

9. (Previously presented) The speech recognition system of claim 7, wherein said human interface device includes a graphical user interface on which the human operator views said text set of expected responses, wherein after listening to said audio response, the human operator is able to select one of said expected responses from said text set of expected responses.

10. (Previously presented) The speech recognition system of claim 9 wherein said graphical user interface comprises an application navigation window for enabling the human operator to navigate through said text set of expected responses, and an audio navigation window for enabling the human operator to control playback of said audio response.

11. (Previously presented) The speech recognition system of claim 8, wherein said graphical user interface comprises an application navigation window for enabling the human

operator to navigate through said text set of expected responses, and an audio navigation window for enabling the human operator to control playback of said audio response.

12. (Previously presented) The speech recognition system of claim 10, wherein said graphical user interface includes a text entry window which enables the human operator to enter a text response if none of said expected responses from said text set of expected responses corresponds to said audio response.

13. (Previously presented) The speech recognition system of claim 9, wherein said graphical user interface includes a text entry window which enables the human operator to enter a text response if none of said expected responses from said text set of expected responses corresponds to said audio response.

14.-25. (Cancelled)

26. (Previously presented) A method of transcribing an audio response comprising:
- A. posing a query to a respondent over the telephone;
 - B. receiving an audio response from said respondent over the telephone;
 - C. performing a speech recognition function on said audio response to automatically convert said audio response to a textual response;
 - D. recording said audio response;
 - E. comparing said textual response to a set of expected responses to said query, said set including a plurality of expected responses to said query in a textual form; and
 - F. flagging said audio response so as to produce a flagged audio response for further review by a human operator if the corresponding textual response does not correspond to one of said expected responses in said set of expected responses within a predetermined accuracy confidence parameter,
 - G. a human operator listening to the actual audio response corresponding to said flagged audio response; and
 - H. a human operator determining if one of said expected responses corresponds to said

actual audio response; and

I. if such determination of step H. is in the affirmative, selecting, from said set of expected responses, a textual response that corresponds to said audio response.

27. (Cancelled)

28. (Previously presented) The method of claim 26, further comprising:

J. manually transcribing a textual response that corresponds to said audio response if such determination of step H is negative.

29. (Previously presented) A method of transcribing an audio response comprising:

A. constructing a speech recognition application including a plurality of queries and a set of expected responses for each query, said set including a plurality of expected responses to each query in a textual form;

B. posing each of said queries to a respondent over the telephone;

C. receiving an audio response to each query from said respondent over the telephone;

D. performing a speaker-independent speech recognition function on each said audio response to automatically convert each said audio response to a textual response to each query;

E. recording and storing each audio response;

F. automatically comparing each textual response to said set of expected responses for each corresponding query to determine if each textual response corresponds to any of said expected responses in said set of expected responses for the corresponding query;

G. flagging an audio response so as to produce a flagged audio response for further review by a human operator if the corresponding textual response does not correspond to one of said expected responses in said set of expected responses within a predetermined accuracy confidence parameter as determined by said speaker-independent speech recognition analysis,

H. a human operator listening to the actual audio response corresponding to said flagged audio response;

I. a human operator determining if one of said expected responses corresponds to said actual audio response; and

J. if such determination of step I. is in the affirmative, the human operator selecting, from said set of expected responses, a textual response that corresponds to said audio response, and flagging each audio response that does not correspond to one of said expected responses in said set of expected responses to the corresponding query.

30.-32. (Cancelled)

33. (Original) The method of claim 29, further comprising manually transcribing a textual response that corresponds to each flagged audio response if such determination of step J is negative.

34.-36 (Cancelled)

REMARKS

As noted previously, the Applicants appreciate the Examiner's thorough examination of the subject application.

Claims 1, 5-13, 26, 28, 29 and 33 remain in the application. Claims 2-4, 14-25, 27, 30-32 and 34-36 have been cancelled. In the Office Action mailed 04 June 2007, claims 1, 5-13, 26, 28, 29, and 33, were rejected as described in further detail below.

Applicants respectfully request reconsideration and further examination of the application based on the foregoing listing of claims and the following remarks.

Claim Rejections – 35 U.S.C. § 103

Concerning items 4-5 of the Office Action, claims 1, 5-13, 26, 28, 29, and 33 were rejected under 35 U.S.C. § 103(a) as being unpatentable over U.S. Patent No. 6,725,194 to Bartosik et al. (“Bartosik”) in view of U.S. Patent No. 6,246,981 to Papineni et al. (“Papineni”). Applicants respectfully traverse this rejection and ask for reconsideration for the following reasons.

One requirement for a rejection under 35 U.S.C. § 103(a) is that the cited reference(s) teach or suggest all of the limitations of the claims at issue. A further requirement for a rejection under 35 U.S.C. § 103(a) is that proper motivation must exist to modify or combine the teachings of the references in the way proposed by the Examiner.

In this situation, the combination of Bartosik and Papineni fails to teach or suggest all of the limitations as recited claim 1 (from which claims 5-13 depend), claim 26 (from which claim 28 depends), and claim 29 (from which claim 33 depends). Furthermore, at least one of the references teaches away from the Examiner's proposed modification/combination, as is explained below.

Representative of the independent claims in the subject application, claim 1 recites the following:

1. A speech recognition system comprising:
a querying device for posing at least one query to a respondent over a telephone;

a speech recognition device which receives an audio response from said respondent over the telephone and conducts a speech recognition analysis of said audio response to automatically produce a corresponding text response;

a storage device for recording and storing said audio response as it is received by said speech recognition device;

an accuracy determination device for automatically comparing said text response to a text set of expected responses and determining whether said text response corresponds to one of said expected responses, wherein said accuracy determination device is configured and arranged to determine whether said text response corresponds to one of said expected responses within a predetermined accuracy confidence parameter and to flag said audio response so as to produce a flagged audio response for further review by a human operator when said text response does not correspond to one of said expected responses within said predetermined accuracy confidence parameter; and

a human interface device for enabling said human operator to hear said flagged audio response and review the corresponding text response for the flagged audio response to determine the actual text response for the flagged audio response, either by selecting from a pre-determined list of text responses or typing the actual text response if no such match exists in the pre-determined list of text responses.

[Emphasis added]

Independent claims 26 and 29 recite method limitations similar to the system limitations recited in claim 1.

In contrast, Bartosik teaches a speech recognition device including speech recognition means arranged for recognizing text information (RTI) corresponding to received voice information (AI) by evaluating the voice information (AI) and a speech coefficient indicator (SKI, PRI, SMI, WI), and including correction means for correcting the recognized text information (RTI) and for producing corrected text information (CTI), and included text comparing means for comparing the recognized text information (RTI) with the corrected text information (CTI) and for determining at least a correspondence indicator (CI) and the adjusting means are provided for adjusting the stored speech coefficient indicator (SKI, PRI, SMI, WI) by evaluating only one of such text parts (P2) of the

corrected text information (CTI) whose correspondence indicator (CI) has a minimum value (MW).
See Bartosik, *e.g.*, Abstract

For the rejection, the Examiner stated that Bartosik teaches, *inter alia*, “an accuracy determination device for automatically comparing said text response to a text set of expected responses and determining whether said text response corresponds to one of said expected responses, wherein said accuracy determination device is configured and arranged to determine whether said text response corresponds to one of said expected responses within a predetermined accuracy confidence parameter and to flag said audio response so as to produce a flagged audio response for further review by a human operator” (e.g., the emphasized portion of claim 1 *supra*), citing Bartosik at col. 6, lines 7-16 and col. 9, lines 1-62.

Applicants disagree as Bartosik is not understood as teaching (or suggesting) flagging of an audio response in the way claimed by Applicants. Bartosik actually teaches systems and methods that functions similar to a dictation machine. *See, e.g.*, Bartosik, col. 3, lines 8-11 (“FIG. 1 shows a computer 1 by which a speech recognition program according to a speech recognition method is run, which computer 1 forms a dictating machine with a secondary speech recognition device.”)

Applicants note that Bartosik relies upon a user reading all recognized text information to determine erroneous recognitions, and because of such actually teaches away from the Applicants’ claims:

The recognized text information RTI recognized by the speech recognition means 42 and stored in the recognized-text memory means 45 is then read out by the text processing means 48 and displayed on the monitor 4. The user recognizes that the two uttered words "order" and "Harry" were recognized erroneously and he/she would like to correct the recognized text information RTI, because of which the user activates with the input means 14 of the dictation microphone 2 the correction mode of the speech recognition device.

(col. 8, lines 6-15) [Emphasis added]

The secondary reference, Papineni, further contrasts with Applicants’ claims by being directed to a speech recognition and synthesis system including a natural language task-oriented dialog manager. For such, Papineni teaches only a general text-to-speech synthesizer. For example, Papineni merely teaches that “hub 10 passes speech data to the speech recognizer 20 which in turns

passes the recognized text back to the hub.” *See* Papineni, col. 7, lines 66-67.

Papineni even goes as far as stating its invention focuses on the dialog manager and script and not the described speech recognizer or text-to-speech synthesizer. *See* Papineni, col. 8, lines 12-18. Papineni clearly does not teach or suggest, *e.g.*, flagging an audio response in the event a predetermined confidence parameter is not met. As a result, Papineni fails to cure the noted deficiencies of Bartosik relative to the Applicants’ claims.

Because of at least the foregoing reasons, the cited combination of Bartosik and Papineni (regardless of whether the references are considered together or separately) is an improper basis for a rejection of claims 1, 5-13, 26, 28, 29, and 33 under 35 U.S.C. § 103(a); Applicants request that the rejection of these claims be removed accordingly.

Response to Arguments

Concerning items 6-8 of the Office Action, the Examiner rebutted the Applicants’ previous remarks concerning the teachings of Bartosik. In response to Applicants’ similar previous assertion concerning Bartosik, the Examiner stated the following:

The examiner disagrees with the applicant’s assertion because as previously examiner stated [sic] that Bartosik teaches, above limitation at col. 6, lines 7-16 and col. 9, lines 1-62 as indicated in the claim rejection. Particularly here claimed “produce a flagged audio response” reads on “in the text comparing means is determined a minimum value MW for the correspondence indicator CI,” “expected response” reads on “possible text information PTI” and claimed “text response” reads on recognized text information RTI”.

Applicants respectfully disagree with the Examiner’s interpretation as Bartosik itself explains a key difference between it and Applicants’ claims: namely, that the systems and methods of Bartosik derive a numerical value (the correspondence indicator CI) that is used for the adjustment of a speech coefficient indicator SKI during operation in a training mode – this correspondence indicator (CI) is not used to flag an audio response in the way claimed by Applicants:

Furthermore, the text comparing means 52, when comparing the recognized text information RTI and the corrected text information CTI, are provided for determining a correspondence indicator CI for each text part. The text comparing means 52 then determine how many matching words featured by a

grey field a text part contains. Furthermore, the text comparing means 52 determine penalty points for each text part, with one penalty point being awarded for each insertion, deletion or substitution of a word in the corrected text information CTI. The correspondence indicator CI of the text part is determined from the number of the corresponding words and penalty points of a text part.

In the text comparing means 52 is determined a minimum value MW for the correspondence indicator CI, which minimum value is fallen short of when for a text part more than three penalty points are awarded for corrections of adjacent words of the corrected text information CTI. For the adjustment of the speech coefficient indicator SKI, only text parts are used whose correspondence indicator CI exceeds the minimum value MW.

(Bartosik, col. 9, lines 43-62) [Emphasis added]

Bartosik further explains that the adjustment of the SKI occurs in a training mode – not a normal use mode:

When the initial training mode is activated, the text processing means 47 are arranged for reading out the training text information TTI from the training-text memory means 47 and for feeding respective picture information PI to the monitor 4. A user can then utter the training text displayed on the monitor 4 into the microphone 6 to adjust the speech recognition device to the user's type of pronunciation [sic].

The speech recognition device has adjusting means 50 for adjusting the speech coefficient indicator SKI stored in the speech-coefficient memory means 38 to the type of pronunciation [sic] of the user and also to words and word sequences commonly used by the user. The text memory means 43, the correction means 49 and the adjusting means 50 together form the training means 51. Such an adjustment of the speech coefficient indicator SKI takes place when the initial training mode is activated in which the training text information TTI read by the user is known.

Such an adjustment, however, also takes place in an adjustment mode in which text information corresponding to voice information is recognized as recognized text information RTI and is corrected by the user into corrected text information CTI. For this purpose, the training means 51 include text comparing means 52, which are arranged for comparing the recognized text information RTI with the corrected text information CTI and for determining at least a correspondence indicator CI. In the text comparing means 52 an adjustment table 53 shown in FIG. 4 is established when the adjustment mode is on, which table will be further explained hereinafter.

(Bartosik, col. 6, line 47 through col. 7, line 9.) [Emphasis added]

Because of such, Applicants submit that Bartosik fails to teach or suggest (and as stated previously actually teaches away from) at least one element of Applicants' claims, *e.g.*, "an accuracy determination device for automatically comparing said text response to a text set of expected responses and determining whether said text response corresponds to one of said expected responses, wherein said accuracy determination device is configured and arranged to determine whether said text response corresponds to one of said expected responses within a predetermined accuracy confidence parameter and to flag said audio response so as to produce a flagged audio response for further review by a human operator when said text response does not correspond to one of said expected responses within said predetermined accuracy confidence parameter;" *e.g.*, as recited in claim 1.

Conclusion


In view of the remarks submitted herein, Applicants respectfully submit that all of the claims now pending in the subject application are in condition for allowance, and therefore request a Notice of Allowance for the application.

Authorization is hereby given to charge any required fees, including those for the Request for Continued Examiner (RCE) under 37 CFR § 1.114 submitted herewith, and to credit any overpayments to deposit account No. 50-1133. If the Examiner believes there are any outstanding issues to be resolved with respect to the above-identified application, the Examiner is invited to telephone the undersigned at his earliest convenience so that such issues may be resolved.

Respectfully submitted,

McDERMOTT WILL & EMERY LLP

Date: 08 August 2007



Toby H. Kusmer, P.C., Reg. No. 26,418
G. Matthew McCloskey, Reg. No. 47,025
28 State Street
Boston, MA 02109
V: (617) 535-4082
F: (617) 535-3800